Predicting Data Quality Success - The Bullwhip Effect in Data Quality

GE, Mouzhi, O'BRIEN, Tony and HELFERT, Markus

Published version

GE, Mouzhi, O'BRIEN, Tony and HELFERT, Markus (2017). Predicting Data Quality Success - The Bullwhip Effect in Data Quality. Perspectives in Business Informatics Research, 157-165.

Predicting Data Quality Success - The Bullwhip Effect in Data Quality

Mouzhi Ge[1], Tony O'Brien[2], Markus Helfert[3]

[1] Faculty of Informatics, Masaryk University, Brno, Czech Republic
Mouzhi.Ge@muni.cz
[2] Sheffield Hallam University, Sheffield, UK
A.O'brien@shu.ac.uk
[3] School of Computing, Dublin City University, Dublin, Ireland
Markus.Helfert@dcu.ie

Abstract. Over the last years many data quality initiatives and suggestions report how to improve and sustain data quality. However, almost all data quality projects and suggestions focus on the assessment and one-time quality improvement, especially, suggestions rarely include how to sustain the continuous data quality improvement. Inspired by the work related to variability in supply chains, also known as the Bullwhip effect, this paper aims to suggest how to sustain data quality improvements and investigate the effects of delays in reporting data quality indicators. Furthermore, we propose that a data quality prediction model can be used as one of countermeasures to reduce the Data Quality Bullwhip Effect. Based on a real-world case study, this paper makes an attempt to show how to reduce this effect. Our results indicate that data quality success is a critical practice, and predicting data quality improvements can be used to decrease the variability of the data quality index in a long run.

Keywords: Data quality, Bullwhip effect, Data quality success, Supply chain, Data quality improvement

1 Introduction

Over the last decade, many researchers and practitioners have emphasised the importance of data quality [3, 4]. Data quality has become a critical concern to the success of organisations. Numerous business initiatives have been delayed or even cancelled, citing poor-quality data as the main reason. Knight and Burn [11] further point out that despite the sizeable body of literature available, relatively few researchers have tackled quantifying the conceptual definitions. The literature provides numerous definitions and taxonomies of data quality dimensions analysing the problem in different contexts [12]. Also, literature provides us with numerous case studies, investigating data quality in practice. Research [7, 9, 14] as well as discussions with practitioners indicate that sustaining and continuously improving data quality in organisations is still challenging. However, most data quality attempts in practice are usually conducted only until the data quality assessment stage or many data quality improvement projects are only executed once.

Continuing the earlier work of [10] in which an approach has been proposed for managing and sustaining data quality in organisation, we expand this work and introduce the concept of Data Quality Bullwhip Effect (DQBE). This paper aims to study the phenomenon and effects of DQBE. It is expected that good data quality prediction models will reduce the observed variability in data quality success indicators, and thus will reduce the DQBE.

In a typical supply chain, consumer's order demand flows up from retailer to wholesaler and in turn goes to distributor, manufacturer and then the raw materials supplier. The bullwhip effect in supply chain occurs when consumer demand indicates a high variability in demand along the levels in the supply chain [13] For example, in order to prevent product shortages and lost sales, the entities in the supply chain tend to order more than they can sell. The extra inventory demand begins from the market fluctuations of supply. When demand increases, the entity like retailer in the lower stream of the supply chain will increase the inventory, likewise, this will amplify the extra inventory on each entity of the supply chain. The Bullwhip effect has been observed across various supply chains in different industries [15].

Following the knowledge from supply chains, the Data Quality Bullwhip Effect is defined as the increase in variability of data quality success over time. Generally, factors that impact the DQBE are related to fuzzy data quality measures and delayed and reluctance to react to the data quality problems. In this paper we considered that the variability of data quality status can be predicted and it helps to reduce the DQBE. We avail of a unique opportunity resulting from a longitudinal case study, in which we observed a data quality programme in a specific organisation. Observing the organisation over time, it revealed variations in the data quality success as well as DQBE. The investigation of this effect and the proposal of forecasting postulates the main contribution of this paper.

The remainder of the paper is organised as follows. Section 2 reviews the related work of data quality success and sustainable data quality improvement. Section 3 describes the Case study of a real-world organisation. Based on the case study, Section 4 discusses the data quality prediction model can how it can be used to prevent bullwhip effect. Finally, section 5 concludes the paper and outline the future research work.

2 Related Work

Reducing the DQBE intends to achieve the data quality success, adapted from Delone and Mclean's IS success model, this paper defines data quality success as sustaining data quality efforts over time in terms of improving data quality. Delone and McLean [2] proposed a comprehensive IS success model that stresses the effects of use and user satisfaction on individual and organizational impact. System quality and data quality are considered as the influencing factors to use and user satisfaction. Importantly, this model brought the attention of data quality to the IS research community and can be considered as one of the pioneering contributions in data quality research. Ten years later, Delone and McLean [1] proposed a revised model that directly focuses on the net benefit instead of organisational impact. As the net benefit can be considered as one form of business value, Delone and McLean' model is not only in line with the model from Gustafsson et al. [8], but also emphasises the importance of data quality and service quality to business value. It can be found that sustainable data quality improvement is an important element in IS success.

DQBE can be observed during continuous data quality improvement. The concept of continuous data quality improvement can be traced back to the 90's. To manage the quality of information products, Wang [16] proposed the total data quality management (TDQM) model to deliver high quality information products. This model is adapted from Deming's plan, do, check and act cycle and consists of four phases: define, measure, analyse and improve. The definition phase is for determining the characteristics of information products, data quality requirements, and how information products are produced in the information manufacturing system. In this phase, we need to identify who assesses the quality of information products and which data quality dimensions are used in such assessment. The measurement phase is for assigning numerical or categorical values to information quality dimensions in a given setting [6]. This phase consists of different measuring methods that can be used to assess data quality. According to the assessment result, the analysis phase is for discovering the root cause of data quality problems and strategizing an effective scheme for data quality improvement. Once the analysis phase is finished, considering budgetary constraints and resource allocation, the improvement phase is concerned with improving the quality of information products for intended use. The four phases constitute a continuous data quality management cycle, indicating that organisations need to continuously implement the TDQM cycle and cultivate data quality concepts into their organisational culture. Although sustainable data quality improvement has been proposed for a long time, there is limited work connecting data quality prediction model and continuous data quality improvement in practice.

Helfert and O'Brien [10] proposed an approach for managing and sustaining data quality in organisation. However, in that work, although the data quality improvement shows significant variabilities, the impact of variations in the data quality success is not investigated. We revisit the important aspects of the case study and use a consolidated case study to show how to reduce the variations in data quality success.

3 Case Study

The case study is related to an UK organisation operating in over sixty individual factories and offices across the entire country. An ERP system was implemented during the latter Nineties and whilst there were many benefits overall it was identified that there was still scope for further improvements especially within the areas of data quality and system complexity. This research study coincided with the commencement of a data quality improvement initiative within the Company during 2005. An initial approach was made across a number of fronts to attempt to promote education and training; documentation of procedures; the acceptance of responsibility, ownership and accountability at all levels for processes and data; together with better management of master data with the identification and implementation of 'quick wins'.

As part of this initiative seven key performance indicators (KPIs) were established around the order fulfilment process, historically the sources of many of the data quality issues. The KPIs were chosen specifically to reflect the salient elements of these essential commercial operations relating to servicing customer needs. The KPIs were designed to reflect the view of the world as seen through the lens of the ERP system, compared with an actual view which could be obtained by direct observation of the actual physical order process. In other words, how closely the 'system' (data within the ERP system) reflects reality (the real world) in the manner described by Wand and Wang [17], whilst also providing a measure of the quality of the actual data and the related processes. From the individual KPIs an aggregated Index was developed weighted to take account of the aging of the various transactions and this was then used as the definitive measurement of the ongoing quality of the data within the KPIs.

3.1 Qualitative Study

The qualitative element of the research took the format of a series of discussion-type focus group meetings sharing experiences, ideas, issues, problems and successes, around a basic flexible agenda, employing an action research approach. This approach attempted to generate discussion and interaction to discover peoples' real feelings and attitudes towards their data. In all, forty-eight of the fifty-four factories and seven business operations and sales teams were covered.

The use of action research in this environment provided the study with a considerable degree of richness in that the researcher had been a member of the organisation for almost twenty years. During this time this 'insider researcher' had worked directly with the majority of the participants and was known to virtually all. This unique approach generated loyalty and trust amongst all parties and not only provided rich material for this study, but also enabled the researcher's colleagues to gain a greater understanding of the significance of quality data and to appreciate the importance of taking ownership of 'their' data. These latter consequences are seen as key to the subsequent improvements that were achieved.

From the outset certain important notions and impressions emerged from the discussions and the analysis and these were subsequently developed as key findings. It was felt that these fell in three broad categories relating to: lessons learnt that should be put in practice at all sites, involving basic quality management principles, ownership, responsibility and support, together with measurement and reporting; positive personal motivational factors which help to engender commitment from individuals, relating to internal competition and targets, an acceptance of best practices and how these relate to one's ideas and principles; together with organizational and cultural environmental elements essentially involving leadership and management issues.

3.2 Measurement and Observation

The importance of measurement, analysis, reporting and feedback was emphasised continually throughout the entire research. Figure 1 below traces the progress of the improvement programme by tracking the Index over the initial three years, highlighting a real trend of improvement over this period, albeit with various explainable fluctuations.

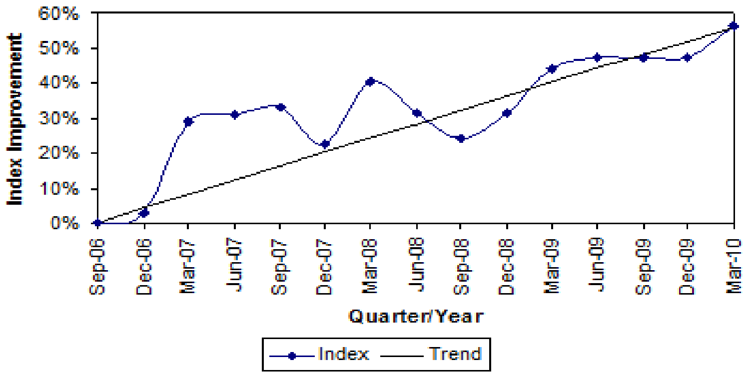

Fig. 1. Data Accuracy KPI Index Improvement tracker

A summary of the progress indicates: 29% improvement in the first six months; 33% improvement in the first year; 16% improvement in the second year; 40% improvement in the first eighteen months; 27% decline in eight months which coincided with the Company's Modernisation Programme; 37% improvement in the year after the modernization; 59% improvement overall within three and a half years. Table 1 below relates the movements in the KPI Index with progression of the site meetings qualitative study both during and following the programme.

Table 1. Data Accuracy KPI Monthly Performance

Month	*11*	*12*	*01*	*02*	*03*	*04*	*05*	*06*	*07*	*08*	*09*	*10*	*11*
Year	Year 1		Year 2										
No of Meetings	0	13	18	3	15	8	0	0	0	0	1	0	0
Index Impr. % Month	0 %	9 %	7 %	-4 %	16 %	1 %	1 %	5 %	0 %	2 %	-3 %	7 %	2 %
Index Impr. % Cum	0 %	9 %	15 %	12 %	26 %	27 %	27 %	31 %	31 %	32 %	30 %	35 %	37 %

It is evident that there was a significant improvement in the Index (27%), following the commencement of the factory and business meeting programme from December to April, in line with the number of meetings carried out. In addition, it may be seen that this level of improvement was maintained immediately following the study and then further improved as the concept of data quality became more established within the organisation.

4 Preliminary Results and Discussion

The direct operational benefits of this Case Study as highlighted by the improved Data Accuracy Index have been referred to in depth, but there is also evidence to suggest that there have also been considerable improvements of a cultural and strategic nature. Further operational and strategic advantages have been derived from enhanced reporting, budgeting and forecasting. The myriad of small meaningful ameliorations, both technical and procedural, which have been applied by passionate people during the period since the original Baan implementation, are now gaining greater maturity alongside higher quality data to generate both operational and informational benefits. Finally, the recognition of the importance of data in relation to overall governance, risk and compliance has provided enhanced levels of authority and control.

Another lessons learned from the case study is how to handle the data quality variability in order to reduce DQBE. Since the data quality variation is featured by the turning points in the prediction curve, we have thus marked some turning points in Figure 2. Each turning point represent a data quality status at a time, for example, point 1 means that in March 2007 the data quality index will be improved by 30%. This curve shows the data quality improvement in every quarter.

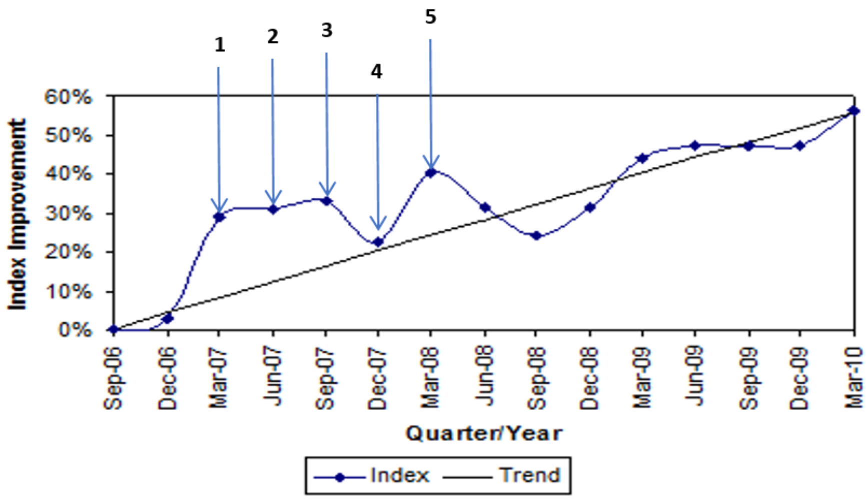

Fig. 2. Data quality turning points marked with numbers

From this case study Figure, we can see that from point 1 to point 3, the data quality improvements are smooth, indicating that there can be limited new data sources or data volumes. In turn, the DQBE can be also limited. However, from point 3 to point 4, there is a significant decrease on the data quality improvement. That means, there might be more data volume or the current data quality improvement cannot handle certain new data source. This fluctuation can possibly cause DQBE. Now the data quality prediction is of significant importance because at point 3 - September 2007, if we know the data quality improvement in the organisation will drop, we could proactively take more data quality improvement actions e.g. shorten the management cycle or include more quality measurements, to prevent the data quality drop. This can then reduce or prevent the DQBE. From point 4 to point 5, if the data quality prediction model indicates that the data quality improvement will be increasing, we could thus release some data quality management resources to reduce the related data quality improvement costs.

Data quality improvement is not just about fixing data or improving quality within a single business application or process, but also about taking a more expansive and forward-looking enterprise-wide approach. This must involve addressing cultural issues, initiating both short and long term process and procedural improvements by a step-by-step, incremental approach, whilst ensuring that the data conforms to appropriate specifications or requirements. In this way any improvement initiative has an opportunity to be sustained. It has to be appreciated that there cannot be a 'one size fits all' remedy to embedding organisational improvements at all levels, but rather to identify appropriate solutions to fit individual situations and circumstances. One accepts that data quality problems are not created intentionally by people, but more by the failure of the surrounding processes whether these are system related or individual related involving lack of education, training, personal developments or purely the person being placed in a position for which they are not suited. There is strong evidence to indicate that solutions exist to improve the quality of data, emanating from both the academic fraternity and the commercial world. This research therefore has not only a strong academic base but also has major practical implications which leads to a further key theme, that of aligning robust theoretical and academic concepts, within the operating environment of a real life organisation, in order to implement sustainable data quality improvements. It is also recognised that research in this specific area may have implications for other functional sectors where process improvements programmes can be applied.

5 Concluding Remarks and Future Research

This research-in-progress paper has conducted a case study from which we have observed the variability in data quality success (index improvement). Inspired from the Bullwhip effect in supply chain, the case study has revealed that there is a similarity between the variability in data quality success and Bullwhip effect in supply chain. Therefore, we termed the variability in data quality success as DQBE. We have used historical data quality improvement as a data quality predication model and discussed that (1) when an organisation should react to which data quality prediction such as increase or decrease. (2) how to proactively react to the data quality prediction in order to reduce the DQBE. Our results further indicate that data quality success is a critical practice, and predicting data quality improvements can be used to decrease the DQBE in a long run.

As future work, we plan to apply the analytical models and frameworks from supply chain research to the data quality domain. As Bullwhip effect is a well-established problem in supply chain, there exists a number of solutions to reduce the Bullwhip effect. These solutions, especially the ones with predication model, may also be used in reducing the data quality bullwhip effect. Another planned work is to validate the DQBE across different contexts in different industries, in which we will simulate the DQBE and use data quality predication model to reduce the DQBE.

One limitation in the paper is that we consider the historical data from the case study as a high quality prediction model, with this model, we can provide the insights to reduce the DQBE. Although the result is limited to the quality of the prediction model, proposing a data prediction model is not in the scope of this paper.

References

1. Delone, W.H., McLean, E.R.: The DeLone and McLean model of information systems success: A ten-year update. Journal of MIS. 19, 9–30 (2003).
2. DeLone, W.H., McLean, E.R.: Information systems success: the quest for the dependent variable. Information systems research. 3, 60–95 (1992).
3. English, L. P. (2009) Information Quality Applied. Indianapolis: Wiley Publications.
4. Gartner (2017) How to Overcome the Top Four Data Quality Practice Challenges. Gartner, Egham UK https://www.gartner.com/doc/reprints?id=1-3W2JDZM&ct=170321&st=sb&elqTrackId=6d44157259df451482de51700cca6d36&elqaid=2250&elqat=2 last access 18.05.2017
5. Ge, M. and Helfert, M., Impact of information quality on supply chain decisions, Journal of Computer Information Systems, 53 (4), 2013
6. Ge, M., Helfert, M., and Jannach, D., 2011 Information Quality Assessment: Validating Measurement Dimensions and Process, in proceedings of 19th European Conference on Information Systems, Helsinki, Finland, 2011.
7. Ge, M. and Helfert, M. (2008), Effects of information quality on inventory management. International Journal of Information Quality, 2(2), pp. 176-191.
8. Gustafsson P., Franke U., Höök D., Johnson P. (2008) Quantifying IT Impacts on Organizational Structure and Business Value with Extended Influence Diagrams. Practice of Enterprise Modeling 2008. vol 15. pp 138-152.
9. Haug, A., Zachariassen, F., & van Liempd, D. (2011) The cost of poor data quality. Journal of Industrial Engineering and Management, 4(2), 168-193.
10. Helfert M., O'Brien T. (2011) Sustaining Data Quality – Creating and Sustaining Data Quality within Diverse Enterprise Resource Planning and Information Systems. Governance and Sustainability in Information Systems. Managing the Transfer and Diffusion of IT. IFIP Advances in Information and Communication Technology, vol 366.
11. Knight, S, Burn, J. (2005) Developing a Framework for Assessing Information Quality on the World Wide Web, Informing Science 8(5)
12. Lukyanenko, R., & Parsons, J. (2015). Information quality research challenge: adapting information quality principles to user-generated content. ACM Journal of Data and Information Quality, 6(1), 3.
13. Lee H.L., Padmanabhan V. and Whang S. (1997) Information distortion in a supply chain: The bullwhip effect, Management Science; Apr 1997; 43, 4; pg. 546

14. O'Brien, T, Sukumar, A and Helfert,M. (2013) The Value of Good Data- A Quality Perspective. International Conference of Enterprise Information Systems, Angers, France
15. Slack, N., Brandon-Jones, A. and Johnston, R. (2016) Operations Management, Eighth edition. Harlow: Pearson Education.
16. Wang, R.Y. (1998), A product perspective on total data quality management, Communications of the ACM, 41(2), pp. 58-65.
17. Wand, Y. and Wang, R.Y. (1996), Anchoring data quality dimensions in ontological foundations, Communications of the ACM, 39(11), pp. 86-95.